NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

TIME SERIES ANALYSIS OF RTC GREAT LAKES RECRUIT GRADUATE DATA

by

Edward F. Bosque

December 1998

Principal Advisor: K.J. Euske
Associate Advisor: Michael Cook

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1998	Master's Thesis

4. TITLE AND SUBTITLE
Time Series Analysis of RTC Great Lakes Recruit Graduate Data

5. FUNDING NUMBERS

6. AUTHOR(S)
Bosque, Edward F..

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

ABSTRACT *(maximum 200 words)*
This thesis formulates predictions for Recruit Training Command (RTC) Great Lakes' recruit graduation rates based on two econometric approaches. The Navy's recruit graduation rates exhibit pronounced seasonal and long-term behaviors, which tends to cause logistical problems at RTC. The modeling and subsequent forecast of RTC graduation rates is therefore an important management tool which could facilitate future planning for both RTC Great Lakes and the US Navy.

First the multiplicative decomposition method is employed to produce a model. As an alternative the autoregressive integrated moving average (ARIMA) process is used to describe the data. In both instances, satisfactory forecasting results are attained.

14. SUBJECT TERMS
Time Series Analysis, Econometrics, Modeling, Recruit Training Command, Recruits

15. NUMBER OF PAGES
53

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFI- CATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited

TIME SERIES ANALYSIS OF RTC GREAT LAKES RECRUIT GRADUATE DATA

Edward F. Bosque
Lieutenant, United States Navy
B.S., University of Colorado, 1993

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGMENT

from the

NAVAL POSTGRADUATE SCHOOL
December 1998

Author: ______________________
Edward F. Bosque

Approved by: ______________________
K.J. Euske, Thesis Advisor

____
Michael Cook, Second Reader

Reuben Harris
____
Reuben Harris, Chairman
Department of Systems Managemet

ABSTRACT

This thesis formulates predictions for Recruit Training Command (RTC) Great Lakes' recruit graduation rates based on two econometric approaches. The Navy's recruit graduation rates exhibit pronounced seasonal and long-term behaviors, which tends to cause logistical problems at RTC. The modeling and subsequent forecast of RTC graduation rates is therefore an important management tool which could facilitate future planning for both RTC Great Lakes and the US Navy.

First the multiplicative decomposition method is employed to produce a model. As an alternative method, we utilize the autoregressive integrated moving average (ARIMA) process to describe the data. In both instances, satisfactory forecasting results area attained.

TABLE OF CONTENTS

I. INTRODUCTION

The Recruit Training Command (RTC) in Great Lakes, Illinois is home to the U.S. Navy's recruit training and the largest training center in the Navy. Since its founding in 1911, RTC has prepared men and women for duty in the naval service. With the closures of RTC Orlando and RTC San Diego in 1994, has been the sole source of recruit training [http://www.ntcpao.com/].

The size of the recruit population at RTC Great Lakes has been and continues to be influenced by external forces. These factors, such as high school graduation dates and the seasons of the year, cause a cyclical inflow of newly reporting personnel that report for basic training. Over sixty percent of the year's accessions arrive between the months of July and November, which causes a number of logistic and ultimately financial difficulties for RTC [Executive Officer, August 1998].

Examples of such difficulties include the placement of Recruit Division Commanders (RDCs) and support staff. While these personnel may be gainfully employed during the peak months, or "surge," the low number of recruits from March to June causes many of the aforementioned RDCs to assume administrative duties or be reassigned to other tasks. Conversely, RDCs are in high demand during the peak months, and staff billets often go unfilled. Other major cost centers affected by this cyclical phenomenon are berthing and messing functions. RTC Great Lakes has only capacity for approximately 1500 recruits at any one time, a constraint imposed by the physical limitations of the base itself [Data Control Officer, September 1998].

A. THESIS OBJECTIVE

The purpose of this thesis is to model the phenomenon of recruit population, or graduation rate. The graduation rate is of particular interest to the Navy, and correlates highly to the accession of new recruit inputs into RTC. Once the graduation rate has been modeled mathematically, it can be used as an accurate predictor of future graduation rates from one to many months in the future. Such knowledge can help RTC Great Lakes and the US Navy in future manpower planning.

B. THESIS ORGANIZATION

This thesis begins with a presentation of the numbers of graduates per month provided by RTC Great Lakes, followed by a time-series analysis of the data. First, the decomposition method is discussed. Then it is employed in an attempt to describe the data. This is followed by the autocorrelation integrated moving average (ARIMA) method, its results, and use in forecasting. Conclusions and recommendations follow.

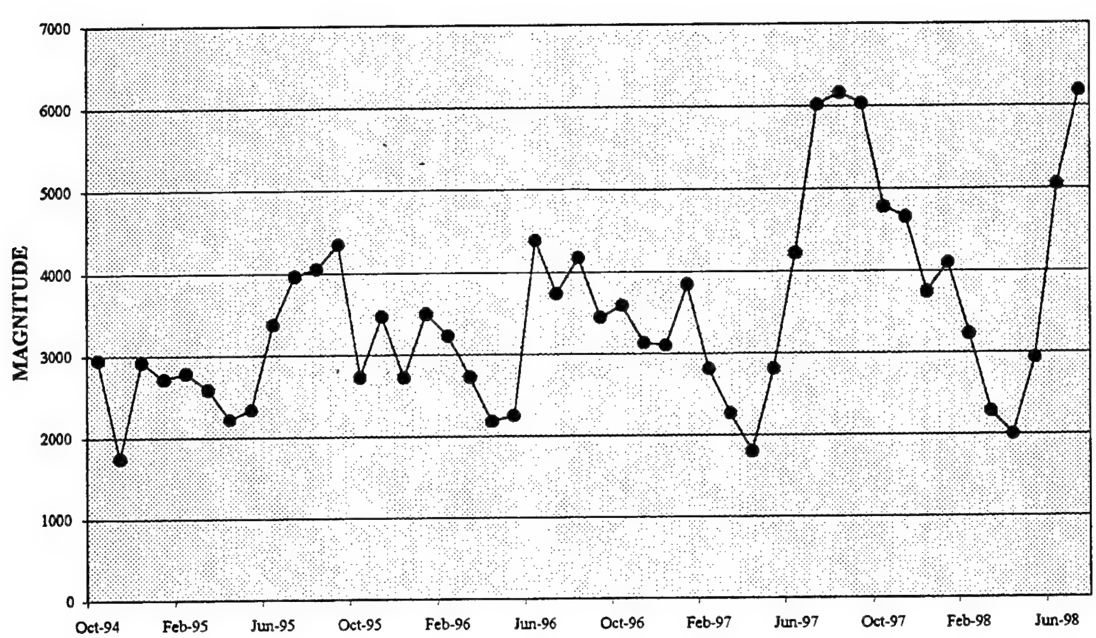
II. DESCRIPTION OF DATA

The data of graduates from RTC Great Lakes starts in October 1994 and concludes July 1998 [Data Control Officer, September 1998]. October 1994 represents the first period in which Great Lakes became the sole source of recruit training [http://www.ntcpao.com/]. Inclusion of previous years' data raises the possibility of inconsistent data, as it does not reflect the total number of recruits.

RTC GREAT LAKES GRADUATION RATES, OCT94 - JUL98					
Period	**Month**	**Graduates**	**Period**	**Month**	**Graduates**
1	Oct-94	2951	25	Oct-96	3588
2	Nov-94	1742	26	Nov-96	3132
3	Dec-94	2923	27	Dec-96	3101
4	Jan-95	2717	28	Jan-97	3843
5	Feb-95	2783	29	Feb-97	2807
6	Mar-95	2585	30	Mar-97	2266
7	Apr-95	2220	31	Apr-97	1798
8	May-95	2336	32	May-97	2806
9	Jun-95	3375	33	Jun-97	4219
10	Jul-95	3962	34	Jul-97	6012
11	Aug-95	4050	35	Aug-97	6159
12	Sep-95	4352	36	Sep-97	6027
13	Oct-95	2727	37	Oct-97	4777
14	Nov-95	3467	38	Nov-97	4655
15	Dec-95	2718	39	Dec-97	3742
16	Jan-96	3497	40	Jan-98	4097
17	Feb-96	3228	41	Feb-98	3230
18	Mar-96	2724	42	Mar-98	2279
19	Apr-96	2178	43	Apr-98	1998
20	May-96	2251	44	May-98	2932
21	Jun-96	4391	45	Jun-98	5042
22	Jul-96	3738	46	Jul-98	6177
23	Aug-96	4176			
24	Sep-96	3445			

The data consists of a series of equally spaced monthly data. This is the underlying definition of a time series, in which the phenomenon in question is a function of time. A graphical presentation of the above data shows the volatility of RTC Great Lakes' graduation rates. The data appears to have a seasonal nature, with a period of approximately twelve months. Also of note is that the data exhibits increased instability, or a more pronounced "seasonal" effect over time. The connecting line between discreet points is for illustrative purposes only.

RTC Great Lakes Graduates OCT94 - JUL98



III. TIME SERIES ANALYSIS

A. DECOMPOSITION METHOD

1. Introduction

A typical time series data set can be considered to be an aggregate of four distinct components. The simplest to understand is the so-called *long-term trend*, which we shall designate **T**. This trend can be negative, positive, or in the case of neither, unchanged. In any event, it may be represented by the linear regression line of a data set or so-called line-of-best-fit. The regression line is calculated as the minimizing the sum of squared errors between a data set and a straight line of the form $y = mx + b$.

Another component is *seasonal variation*. This behavior is typified by a data set's change in values according to the time of year or other seasonal regularity such as the weather. Seasonal variation, or **S**, is repetitive in nature and very similar to *cyclical variation*, **C**. The distinction between seasonal and cyclical variation lies in the fact that seasonal variation has specific fixed time intervals, and cyclical variation does not. Cyclical variation can last any specified length of time, which is sometimes regarded as a business cycle. Also of consequence in all time series analysis is *random variation*, **R**. Random variation can account for the lack of any identifiable data pattern, and is almost always present to some extent in any real set of data points [Gujurari, 1995].

Time series data can be viewed as a combination of the above behaviors [Gujarati, 1995]. Mathematically speaking, if we consider the variable **Y** to be the phenomenon under observation, **Y** may be expressed as the product of the four aforementioned behavior patterns:

$$Y = T \cdot S \cdot C \cdot R \qquad (1)$$

Where

T = long-term trend,
S = seasonal variation,
C = cyclical variation, and
R = random variation.

This model captures all of the aforementioned behavior patterns. Since Equation (1) is a multiplicative model, these components are superimposed on each other, forming an aggregate pattern. Equation (1) allows each component to be manipulated or isolated [Gurarati, 1995].

2. Moving Averages

Paramount to the decomposition method is the calculation of the data's *moving average*, **MA**. To obtain accurate figures, we shall use a *centered moving average* which is centered to the middle of the data points in question. Since we are using monthly data, we will employ a twelve-period centered moving average of the form

$$MA_i = \frac{Y_{i-6} + 2 \cdot \Sigma\,(Y_{i-5} + Y_{i-4} + \ldots + Y_{i} + \ldots + Y_{i+4}) + Y_{i+5}}{22} \qquad (2)$$

By employing spreadsheets, RTC Great Lakes' moving averages for graduate data, October 1994 to July 1998 is calculated as follows:

	RTC GRADUATES, OCT94 - JUL98						
Period	Month	Graduates	Moving Avg	Period	Month	Graduates	Moving Avg
		Y	MA			Y	MA
1	Oct-94	2951		25	Oct-96	3588	3058
2	Nov-94	1742		26	Nov-96	3132	3043
3	Dec-94	2923		27	Dec-96	3101	2958
4	Jan-95	2717		28	Jan-97	3843	2912
5	Feb-95	2783		29	Feb-97	2807	3008
6	Mar-95	2585		30	Mar-97	2266	3198
7	Apr-95	2220	2695	31	Apr-97	1798	3413
8	May-95	2336	2795	32	May-97	2806	3583
9	Jun-95	3375	2859	33	Jun-97	4219	3716
10	Jul-95	3962	2881	34	Jul-97	6012	3777
11	Aug-95	4050	2911	35	Aug-97	6159	3826
12	Sep-95	4352	2968	36	Sep-97	6027	3920
13	Oct-95	2727	3015	37	Oct-97	4777	3980
14	Nov-95	3467	3030	38	Nov-97	4655	3967
15	Dec-95	2718	2976	39	Dec-97	3742	3879
16	Jan-96	3497	2947	40	Jan-98	4097	3785
17	Feb-96	3228	2952	41	Feb-98	3230	3746
18	Mar-96	2724	2932	42	Mar-98	2279	
19	Apr-96	2178	2955	43	Apr-98	1998	
20	May-96	2251	2989	44	May-98	2932	
21	Jun-96	4391	3012	45	Jun-98	5042	
22	Jul-96	3738	3013	46	Jul-98	6177	
23	Aug-96	4176	3022				
24	Sep-96	3445	3051				

The use of moving averages smoothes short-term fluctuations by averaging any data point that may be unusually high or low [Judge, et all, 1985]. Since each period covers a complete cycle of observation, in our case twelve months, the data's moving average can be considered a product of its long-term trend and cyclical variance [Gujarati, 1995]:

$$MA = T \cdot C \tag{3}$$

By incorporating Equation (1), Equation (3) becomes

$$Y = MA \cdot S \cdot R, \text{ or}$$

$$Y/MA = S \cdot R \tag{4}$$

The ratio Y/MA is called the actual-to-moving average ratio. It is an important relationship as seasonal and random variances can be isolated [Gujurati, 1995]. Said another way, the actual-to-moving ratio is said to contain seasonality and randomness.

3. Seasonality

We can now de-seasonalize. This is done by averaging all moving-to-average ratios found previously by month for all years to obtain *seasonal indices*, **S**. Each seasonal index corresponds to a specific month, and is found in the last column :

Y/MA TABLE -- DETERMINATION OF SEASONAL INDICES						
					Arithmetic	Adjusted
	1995	1996	1997	1998	Mean	Indices
Jan		1.1865	1.3198	1.0824	1.1962	1.0918
Feb		1.0934	0.9331	0.8623	0.9630	0.8789
Mar		0.9291	0.7086		0.8188	0.7474
Apr	0.8236	0.7372	0.5269		0.6959	0.6351
May	0.8358	0.7530	0.7832		0.7907	0.7216
Jun	1.1807	1.4580	1.1354		1.2580	1.1482
Jul	1.3751	1.2408	1.5919		1.4026	1.2802
Aug	1.3913	1.3820	1.6097		1.4610	1.3335
Sept	1.4665	1.1292	1.5375		1.3778	1.2575
Oct	0.9044	1.1734	1.2002		1.0926	0.9973
Nov	1.1443	1.0294	1.1736		1.1157	1.0183
Dec	0.9132	1.0484	0.9646		0.9754	0.8903
Sum					13.1478	12.0000

As the sum of the arithmetic mean does not add to 12, the indices can be adjusted by multiplying each average by the quotient of 12.0000/13.1478. The sum of indices should

add to twelve, corresponding to the number of months in a year. We next obtain the de-seasonalized data. Dividing both sides of Equation (1) by S, we obtain the equation:

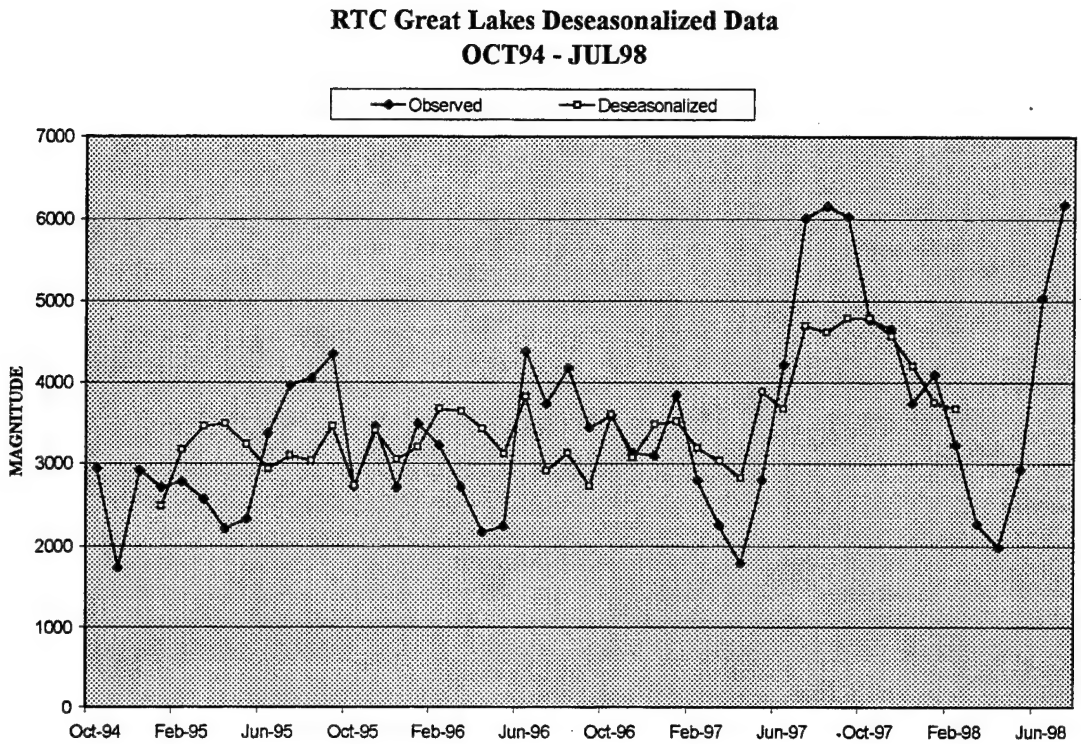
$$Y/S = T \cdot C \cdot R \qquad \text{........(5)}$$

Our original data now take the form:

	RTC GRADUATES, OCT94 - JUL98					
					Adjusted	**Deseasonalized**
Period	**Month**	**Graduates**	**Moving Avg**		**Seasonal Index**	**Data**
		Y	**MA**	**Y/MA**	**S**	**Y/S**
1	Oct-94	2951				
2	Nov-94	1742				
3	Dec-94	2923				
4	Jan-95	2717			1.0918	2489
5	Feb-95	2783			0.8789	3167
6	Mar-95	2585			0.7474	3459
7	Apr-95	2220	2695	0.8236	0.6351	3495
8	May-95	2336	2795	0.8358	0.7216	3237
9	Jun-95	3375	2859	1.1807	1.1482	2939
10	Jul-95	3962	2881	1.3751	1.2802	3095
11	Aug-95	4050	2911	1.3913	1.3335	3037
12	Sep-95	4352	2968	1.4665	1.2575	3461
13	Oct-95	2727	3015	0.9044	0.9973	2734
14	Nov-95	3467	3030	1.1443	1.0183	3405
15	Dec-95	2718	2976	0.9132	0.8903	3053
16	Jan-96	3497	2947	1.1865	1.0918	3203
17	Feb-96	3228	2952	1.0934	0.8789	3673
18	Mar-96	2724	2932	0.9291	0.7474	3645
19	Apr-96	2178	2955	0.7372	0.6351	3429
20	May-96	2251	2989	0.7530	0.7216	3119
21	Jun-96	4391	3012	1.4580	1.1482	3824
22	Jul-96	3738	3013	1.2408	1.2802	2920
23	Aug-96	4176	3022	1.3820	1.3335	3132
24	Sep-96	3445	3051	1.1292	1.2575	2740

	RTC GRADUATES, OCT94 - JUL98 (Continued)					
					Adjusted	Deseasonalized
Period	Month	Graduates	Moving Avg		Seasonal Index	Data
		Y	MA	Y/MA	S	Y/S
25	Oct-96	3588	3058	1.1734	0.9973	3598
26	Nov-96	3132	3043	1.0294	1.0183	3076
27	Dec-96	3101	2958	1.0484	0.8903	3483
28	Jan-97	3843	2912	1.3198	1.0918	3520
29	Feb-97	2807	3008	0.9331	0.8789	3194
30	Mar-97	2266	3198	0.7086	0.7474	3032
31	Apr-97	1798	3413	0.5269	0.6351	2831
32	May-97	2806	3583	0.7832	0.7216	3888
33	Jun-97	4219	3716	1.1354	1.1482	3675
34	Jul-97	6012	3777	1.5919	1.2802	4696
35	Aug-97	6159	3826	1.6097	1.3335	4619
36	Sep-97	6027	3920	1.5375	1.2575	4793
37	Oct-97	4777	3980	1.2002	0.9973	4790
38	Nov-97	4655	3967	1.1736	1.0183	4571
39	Dec-97	3742	3879	0.9646	0.8903	4203
40	Jan-98	4097	3785	1.0824	1.0918	3753
41	Feb-98	3230	3746	0.8623	0.8789	3675
42	Mar-98	2279				
43	Apr-98	1998				
44	May-98	2932				
45	Jun-98	5042				
46	Jul-98	6177				

These de-seasonalized values can be represented graphically:



The estimate of the trend line **T** is found by using the de-seasonalized data. To proceed further, the resultant least-squares equation for **T** can be found by means of a simple linear regression calculated by the MINITAB© release 12.1 software package:

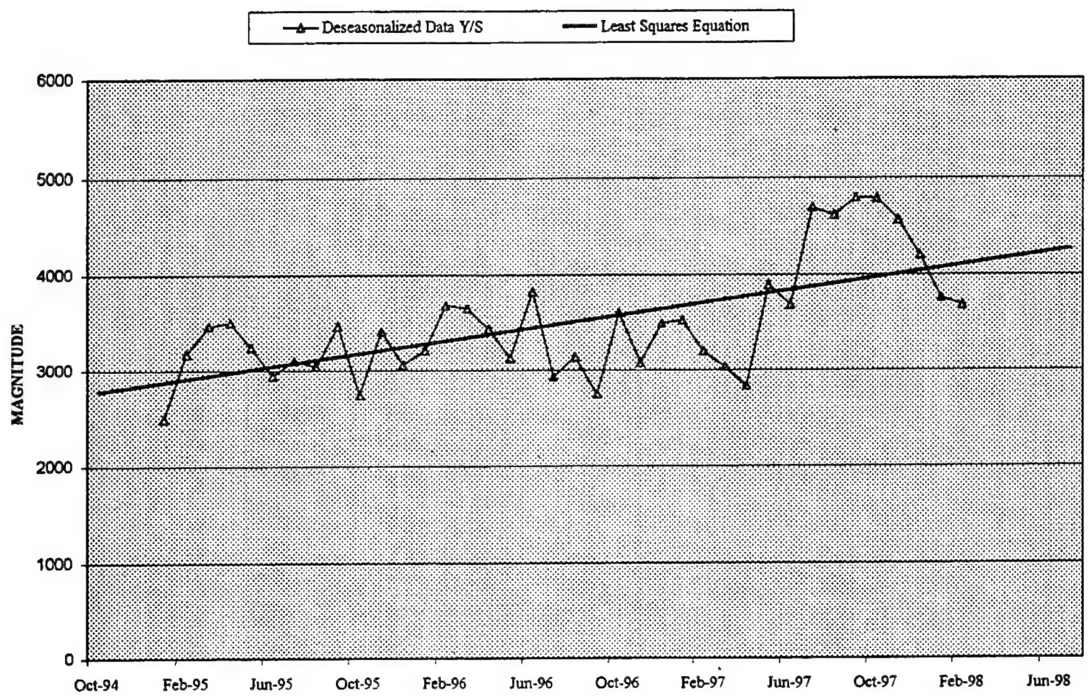
Regression Analysis

```
The regression equation is
Y/S = 2743 + 33.2 Period

Predictor        Coef       StDev          T        P
Constant       2743.4       174.3      15.74    0.000
Period         33.223       6.964       4.77    0.000
```

The graph and table of de-seasonalized data and the least squares equation values follow. The trend line values follow readily from the least squares equation and are computed using a spreadsheet.

RTC Great Lakes Deseasonalized Data OCT94 - JUL98



	RTC GRADUATES, OCT94 - JUL98				
Period	Deseasonalized Data Y/S	Least Squares Equation T	Period	Deseasonalized Data Y/S	Least Squares Equation T
1		2776	25	3705	3573
2		2809	26	3484	3606
3		2843	27	3775	3639
4	1855	2876	28	2624	3673
5	2361	2909	29	2381	3706
6	2579	2942	30	2261	3739
7	4304	2975	31	3486	3772
8	3727	3009	32	4476	3805
9	3189	3042	33	3987	3839
10	3428	3075	34	5201	3872
11	3318	3108	35	5045	3905
12	3999	3141	36	5539	3938
13	2816	3175	37	4932	3971
14	3857	3208	38	5179	4005
15	3309	3241	39	4556	4038
16	2388	3274	40	2798	4071
17	2738	3307	41	2740	4104
18	2718	3341	42		4137
19	4223	3374	43		4171
20	3591	3407	44		4204
21	4150	3440	45		4237
22	3234	3473	46		4270
23	3421	3507			
24	3166	3540			

4. Forecasting

Having obtained the adjusted seasonal indices and trend line values, we can construct a forecasting model of the form [Gujurati, 1995]

$$\hat{Y} = S \cdot T \text{, or} \qquad \text{........(6)}$$

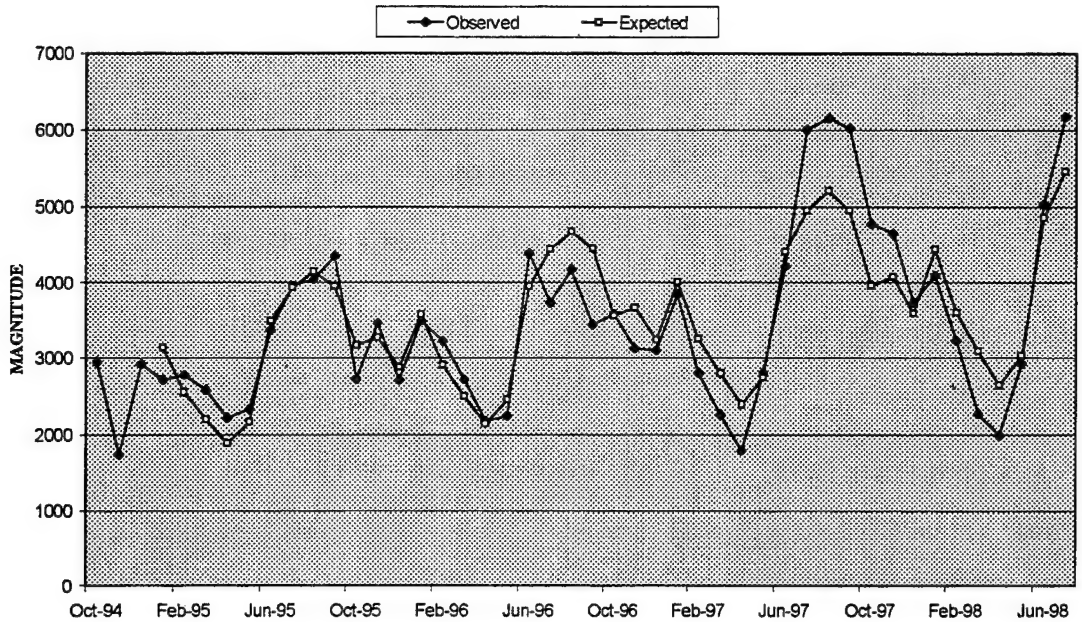
$$\hat{Y} = S \cdot (2743 + 33.2 \cdot \text{Period})$$

The consequent values from Equation (6) are as follows:

RTC GRADUATES, OCT94 - JUL98							
		Seasonal	Forecast			Seasonal	Forecast
Period	Graduates	Index	Value	Period	Graduates	Index	Value
	Y	S			Y	S	
1	2951			27	3101	0.8880	3232
2	1742			28	3843	1.1824	4342
3	2923			29	2807	1.0285	3811
4	2717	1.1824	3400	30	2266	1.2188	4557
5	2783	1.0285	2992	31	1798	1.0569	3987
6	2585	1.2188	3586	32	2806	0.5860	2230
7	2220	1.0569	3145	33	4219	0.9323	3579
8	2336	0.5860	1763	34	6012	1.0395	4025
9	3375	0.9323	2836	35	6159	1.1147	4353
10	3962	1.0395	3197	36	6027	1.0864	4278
11	4050	1.1147	3465	37	4777	0.9147	3633
12	4352	1.0864	3413	38	4655	0.9517	3811
13	2727	0.9147	2904	39	3742	0.8880	3585
14	3467	0.9517	3053	40	4097	1.1824	4813
15	2718	0.8880	2878	41	3230	1.0285	4221
16	3497	1.1824	3871	42	2279	1.2188	5043
17	3228	1.0285	3402	43	1998	1.0569	4408
18	2724	1.2188	4072	44	2932	0.5860	2463
19	2178	1.0569	3566	45	5042	0.9323	3950
20	2251	0.5860	1996	46	6177	1.0395	4439
21	4391	0.9323	3207				
22	3738	1.0395	3611				
23	4176	1.1147	3909				
24	3445	1.0864	3846				
25	3588	0.9147	3268				
26	3132	0.9517	3432				

The resultant values of Equation (6) and actual observed values are represented graphically:

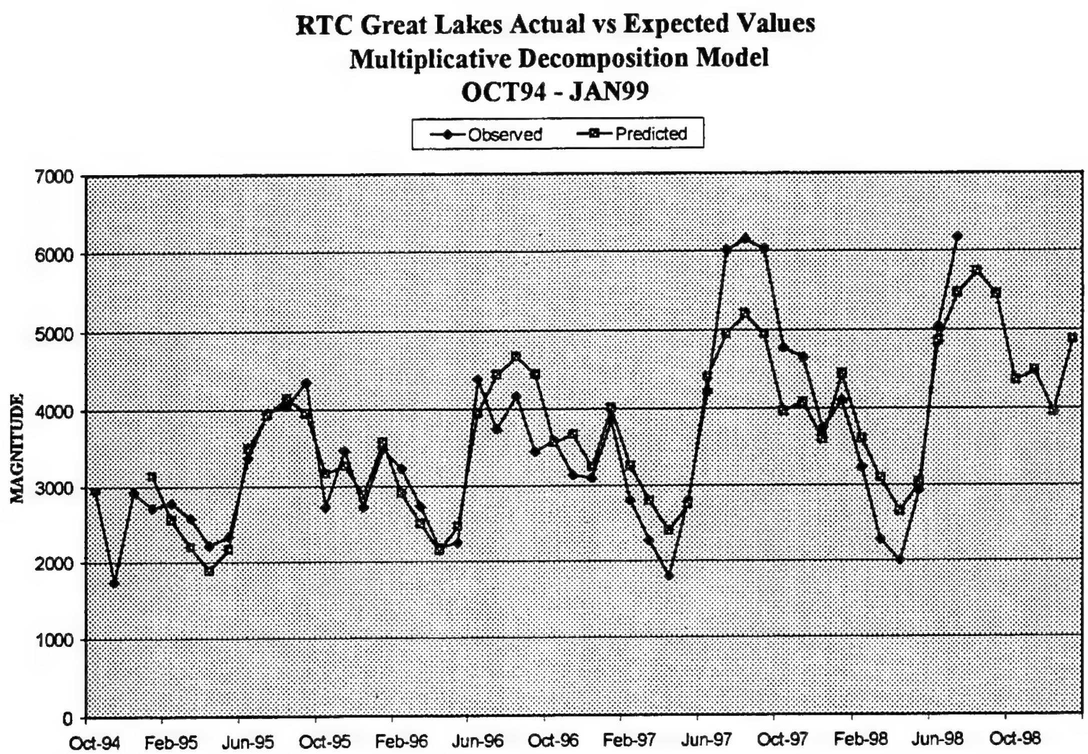
RTC Great Lakes Actual vs Expected Values
Multiplicative Decompoition Method
OCT94 - JUL98



We can extend Equation (6) to establish a forecast of future RTC Great Lakes graduation rates for the multiplicative decomposition method:

RTC GREAT LAKES GRADUATE FORECAST				
Period	Month			Forecast Value
47	Aug-98			4797
48	Sep-98			4711
49	Oct-98			3997
50	Nov-98			4190
51	Dec-98			3939
52	Jan-99			5285

A graph of actual and expected values including forecast figures, appears as follows:

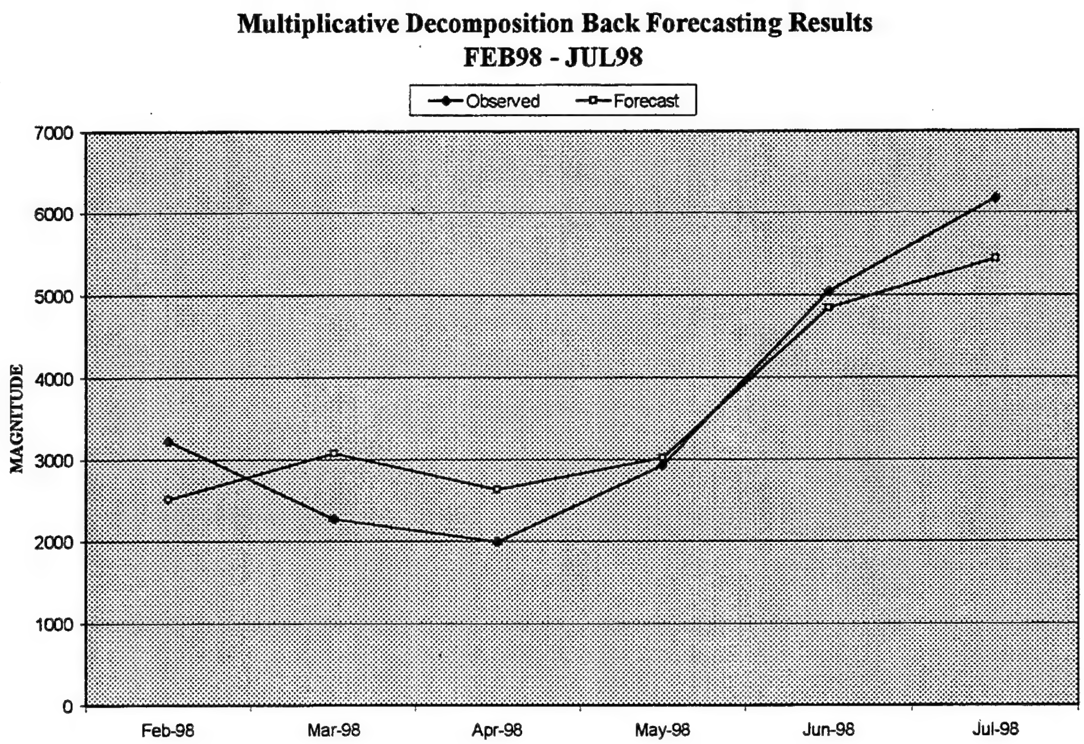


5. Forecast Error

The underlying assumption in any time series forecast is that the time series will behave in the future as it did in the past. A point forecast, which corresponds to a discreet data point, represents the best prediction of the value of the variable in question at any given point in the future. It is our "best guess" for the future value of the variable [Harvey, 1993].

In order to ascertain the validity of the decomposition model we performed a similar analysis, this time with only forty observed values. This separate analysis included data from October 1994 through January 1998. As expected, different moving averages and

seasonal indices were produced. A forecast was then executed for the remaining six observed periods, February 1998 through July 1998, and that forecast was contrasted to the actual observed values of the data from that time period. This procedure is known as a *back forecast*, and is utilized as a preliminary litmus test of the model under review [Box and Jenkins, 1970]



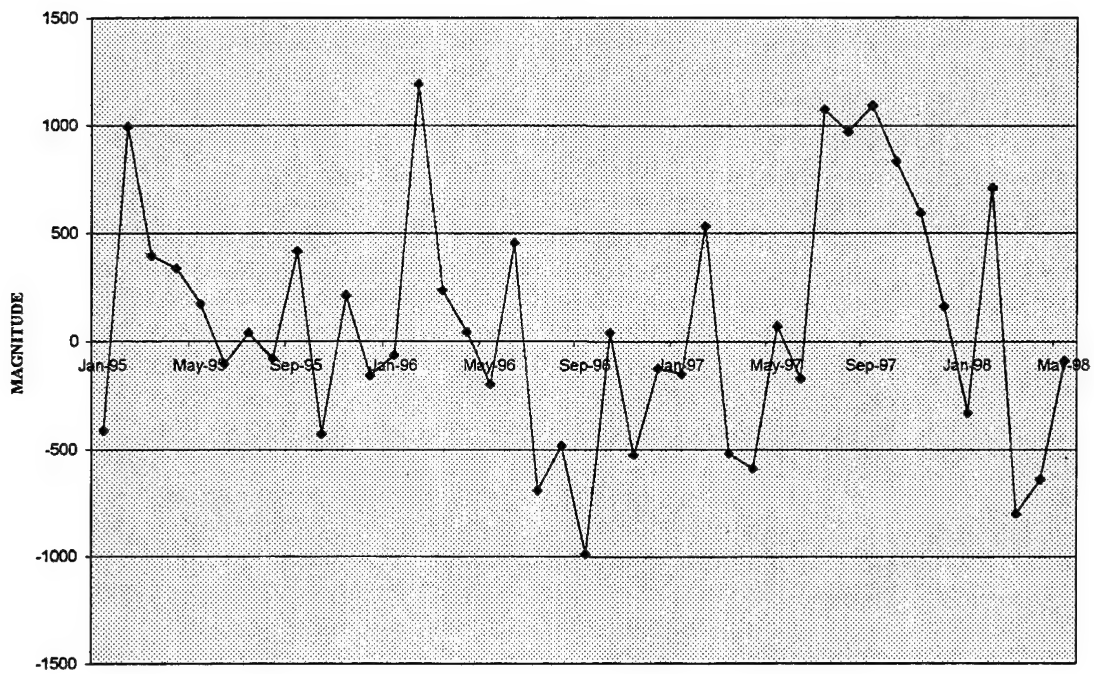
A visual inspection of the back forecasting provides an indication that the multiplicative decomposition model is appropriate. The back forecasting results appear to model the actual observations. The degree of appropriateness, or amount of quantifiable error inherent in our model shall be discussed shortly.

Unfortunately, all attempts at forecasting involves some degree of uncertainty which increases the further one is removed from the origin of the forecast, period t [Box and

Jenkins, 1970]. Unpredictable fluctuations inherent in the data imply that some error in forecasting must be expected. A large degree of variance σ^2 in these fluctuations will limit the accuracy of our forecasts [Bowerman and O'Connell, 1979]. Conversely, a smaller variance of the irregular component of the data will allow us to forecast with greater confidence in the results. Another aspect of forecast error comes from the type and specifications of the forecast model itself. The accuracy with which we derive or select the components of the time series model influences the error inherent in our forecast [Bowerman and O'Connell, 1979]. The better the model describes the data, the less the degree of forecasting error.

An examination of forecast errors over a large period of time can reveal whether the forecasting technique used is relevant. In the case of decomposition, we should expect that all seasonal, trend, or cyclical components of the data have been eliminated, leaving only a random component [Bowerman and O'Connell, 1979]. This can be seen graphically with the residual plot below. Residual values represent the difference between observed values and those values predicted by the forecast equation. For a forecast method to be accurate, its residual plot should exhibit no discernable pattern. In the following data we have not identified a distinct pattern over time:

Multiplicative Decomposition Residual Plot



Quantifying the error of the model is a straightforward procedure. We use the mean squared error (MSE) of the forecasts, which is simply the average of the squared errors for all forecasts. The following table shows the MSE for the forecasts of periods forty-one through forty-six, the back forecasted data:

MULTIPLICATIVE DECOMPOSITION MSE CALCULATION						
		Observed	Forecast		Squared	
Period	Month	Value	Value	Error	Error	MSE
		Z_t	Z_e	$e_t = Z_t - Z_e$		
41	Feb-98	3230	2521	709	503,084.38	356,170.14
42	Mar-98	2279	3080	-801	642,141.74	
43	Apr-98	1998	2639	-641	410,676.87	
44	May-98	2932	3022	-90	8,118.69	
45	Jun-98	5042	4846	196	38,304.43	
46	Jul-98	6177	5446	731	534,694.73	
				Sum	2,137,020.83	

By itself, the MSE figure for the multiplicative decomposition method tells us little. When compared to other forecast methods' MSE, however, it can be used to aid the process of forecast technique or model selection [Bowerman and O'Connell, 1979] with lower MSE scores being preferable [Kennedy, 1979]. We shall compare the multiplicative decomposition method's MSE with another model shortly. For now we can assume by way of the back forecast and residual plot that the multiplicative decomposition method is in itself a relevant model which can be used to adequately forecast RTC Great Lakes' future graduation rates.

B. AUTOCORRELATION INTEGRATED MOVING AVERAGE METHOD

1. Introduction

The use of mathematical models to describe the behavior of a particular phenomenon has been thoroughly established [Harvey, 1993]. One might use equations to calculate an object's trajectory through space or pH levels in a chemical process. No process is entirely deterministic, however, as unknown factors tend to wreak havoc with deterministic models and equations. It is important to recognize that *randomness* is always present to some extent in a data set. Deterministic models lack the ability to quantify or codify outside forces into a coherent mathematical expression. For example, an investor can know virtually everything about a corporation and possess the latest macroeconomic data, however, accurately forecasting the corporation's stock price on a daily basis is for all practical purposes impossible.

While it may prove futile to write a deterministic model which exactly calculates the future behavior of a probabilistic process, it may be possible to derive an expression which models data within specified limits [Harvey, 1993]. Such a probabilistic process is also referred to as a *stochastic* process. A stochastic model defines a mechanism which is regarded as being capable of generating the observed values in question [Harvey 1993].

2. Mathematical Terms and Expressions

a. Indices

t	time
$t+l$	future time l units distant

b. Operators

B	backward-shift operator
∇	backward-difference operator
$\Phi(B)$	autoregressive operator
$\vartheta(B)$	generalized nonstationary autoregressive operator
$\Theta(B)$	moving average operator

c. Data

Z_t	graduates in current month t

d. Variables

Φ_p	autoregressive variable of order p
ϑ_j	generalized nonstationary autoregressive variable of order j
Θ_q	moving average variable of order q
a_t	shock or noise at time t
$\hat{Z}_t$	deviation from trend μ at time t ($\hat{Z}_t = Z_t - \mu$)
$\tilde{Z}_t(l)$	forecast made at origin t of the graduates Z_{t+l} at future time t+l

The backward-shift operator B is defined

$$BZ_t = Z_{t-1}$$

More generally, we can say

$$B^m Z_t = Z_{t-m}$$

The backward-difference operator ∇ is defined

$$\nabla Z_t = Z_t - Z_{t-1} = (1 - B)Z_t$$

3. Autoregressive Processes

A stochastic model that has proven to be particularly useful in the representation of certain time series data is the *autoregressive model* [Box and Jenkins, 1970]. In this model, the current value of the process Z_t is expressed as a finite, linear aggregate of the process' previous values, $Z_{t-1}, Z_{t-2,...,} Z_{t-T}$, and noise, a_t. If we let μ represent the level of the process, we write a first-order autoregressive process, designated AR(1) [Box and Jenkins, 1970],

$$\hat{Z}_t = \Phi_1 \hat{Z}_{t-1} + a_t, \quad t = 1...T \tag{7}$$

In the case of AR(1), the model depends only on the previous value of the data. Likewise, the second-order autoregressive process, AR(2) is defined by

$$\hat{Z}_t = \Phi_1 \hat{Z}_{t+1} + \Phi_2 \hat{Z}_{t+2} + a_t, \quad t = 1...T \tag{8}$$

In general, we may write an expression for an autoregressive process of order p:

$$\hat{Z}_t = \Phi_1 \hat{Z}_{t+1} + \Phi_2 \hat{Z}_{t+2} + \Phi_3 \hat{Z}_{t+3} + ... + \Phi_{p1} \hat{Z}_{t+p} + a_t, \quad t = 1...T \tag{9}$$

It is possible to determine the appropriateness of the autoregressive process to the time series in question by means of the data's autocorrelation graph (ACF). Autocorrelation describes the mutual dependence among values of the same variable Z_t at different periods. If the data set contains purely random values, the autocorrelation

among successive values will be close to or equal to 0. Conversely, data that exhibits a definite dependence on previous values of the variable Z_t will be highly correlated [Box and Jenkins, 1970]. The data plot which follows illustrates the ACF graph for RTC Great Lakes' graduation data. The gradual decrease of the autocorrelation coefficients, as opposed to a sudden drop to 0, suggests the appropriateness of the AR model in the case of the RTC Great Lakes data [Box and Jenkins, 1970]. The ACF graph is the product of the MINITAB© software package, release 12.1.

Autocorrelation Function

ACF of RTC Great Lakes Graduates

```
            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
              +----+----+----+----+----+----+----+----+----+----+
  1  0.643                             XXXXXXXXXXXXXXXXX
  2  0.274                             XXXXXXXX
  3 -0.060                           XXX
  4 -0.138                          XXXX
  5 -0.122                          XXXX
  6 -0.141                         XXXXX
  7 -0.180                        XXXXXX
  8 -0.215                        XXXXXX
  9 -0.064                           XXX
 10  0.154                             XXXXX
 11  0.406                             XXXXXXXXXXX
 12  0.416                             XXXXXXXXXXX
 13  0.242                             XXXXXXX
 14 -0.009                             X
 15 -0.146                         XXXXX
 16 -0.128                          XXXX
 17 -0.067                           XXX
 18 -0.061                           XXX
 19 -0.170                         XXXXX
 20 -0.172                         XXXXX
 21 -0.124                          XXXX
 22  0.039                             XX
 23  0.180                             XXXXX
 24  0.240                             XXXXXXX
 25  0.141                             XXXXX
 26 -0.052                            XX
 27 -0.107                          XXXX
 28 -0.086                           XXX
 29 -0.078                           XXX
 30 -0.139                          XXXX
 31 -0.216                        XXXXXX
 32 -0.238                       XXXXXXX
 33 -0.192                        XXXXXX
 34 -0.056                            XX
 35  0.015                             X
```

4. Moving Average Processes

Another model, the so-called *moving average* process, expresses Z_t as a finite number q of current and previous shocks in the system, $a_t, a_{t-1}, \ldots, a_{t-q}$. This is referred to as a moving average (MA) process of order q, or MA(q). The general form of the process is written [Box and Jenkins, 1970]:

$$\hat{Z}_t = a_t - \Theta_1 a_{t-1} + \Theta_2 a_{t-2} + \Theta_3 a_{t-3} + \ldots + \Theta_q a_{t-q}, \quad t = 1 \ldots T \tag{10}$$

Moving average models imply that what occurs at time t is not influenced by previously observed values of the variable in question, nor will it be influenced by future events. It is also referred to as the White Noise model [Box and Jenkins, 1970].

Like its counterpart the AR process, the appropriateness of the MA process to a particular data set may be determined by means of a graph, in this case the partial autocorrelation coefficient (PACF) plot [Box and Jenkins, 1970]. The PACF for RTC Great Lakes graduation data follows. Note the coefficients' gradual reduction. This data plot is also generated by the MINITAB© software package.

Partial Autocorrelation Function

PACF of RTC Great Lakes Graduates

```
          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
  1   0.643                          XXXXXXXXXXXXXXXXX
  2  -0.239                    XXXXXXX
  3  -0.226                      XXXXX
  4   0.129                          XXXX
  5  -0.025                         XX
  6  -0.191                     XXXXXX
  7  -0.063                        XXX
  8  -0.051                         XX
  9   0.202                          XXXXX
 10   0.157                          XXXXX
 11   0.235                          XXXXXXX
 12  -0.040                         XX
 13  -0.107                       XXXX
 14  -0.087                        XXX
 15  -0.005                          X
 16   0.055                          XX
 17   0.057                          XX
 18  -0.016                          X
 19  -0.135                       XXXX
 20   0.032                          XX
 21  -0.118                       XXXX
 22  -0.034                         XX
 23   0.101                          XXXX
 24   0.131                          XXXX
 25  -0.044                         XX
 26  -0.160                      XXXXX
 27   0.052                          XX
 28  -0.018                          X
 29  -0.190                     XXXXXX
 30   0.021                          XX
 31   0.034                          XX
 32  -0.044                         XX
 33  -0.137                       XXXX
 34  -0.019                          X
 35  -0.156                      XXXXX
 36  -0.091                        XXX
```

5. Mixed Autoregressive-Moving Average Models

To obtain greater flexibility in modeling time series data, it is usually advantageous to include both autoregressive (AR) and moving average (MA) terms in the stochastic model [Box and Jenkins, 1970]. Combining Equation (9) and Equation (10) provides a mixed process of AR and MA elements known as an autoregressive-moving average process of order (p,q) or ARMA(p,q):

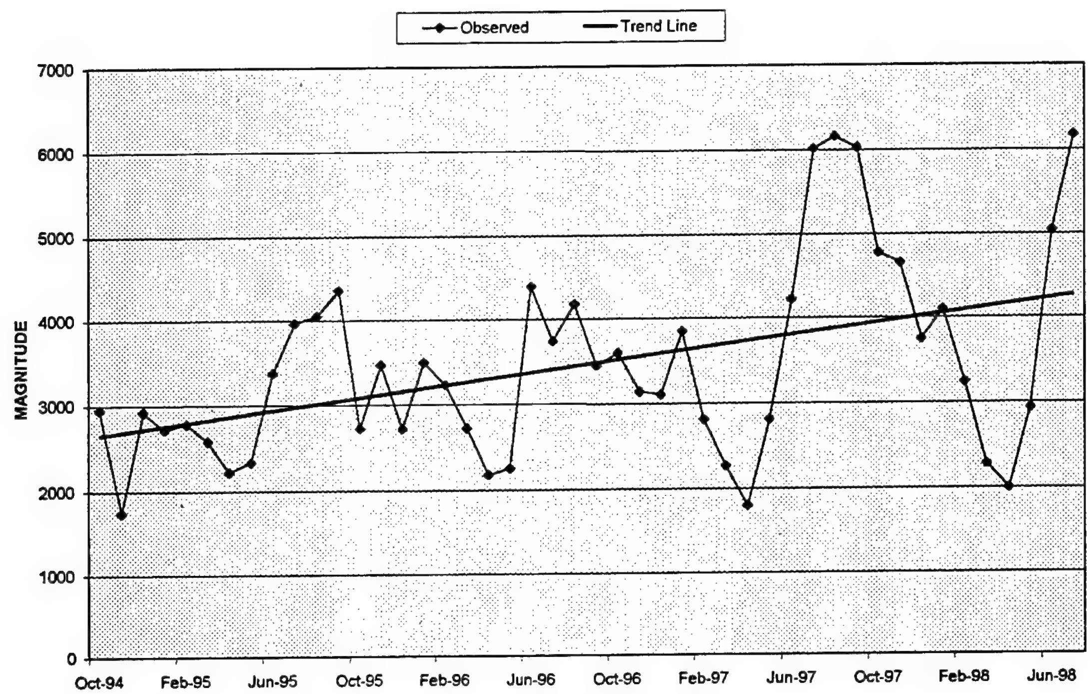
$$\hat{Z}_t = \Phi_1 \hat{Z}_{t-1} + \ldots + \Phi_{p1} \hat{Z}_{t-p} + a_t - \Theta_1 a_{t-1} - \ldots - \Theta_q a_{t-q}, \quad t = 1..T \qquad (11)$$

Since $\nabla^d \hat{Z}_t = \nabla^d Z_t$ for d $\geq$ 1, we can replace $\hat{Z}_t$ with Z_t [Box and Jenkins, 1970]. Equation (11) portrays the dependent variable not only as a function of previous observations, but also previous deviations caused by ambient noise. This non-linear equation is highly effective in modeling a wide array of behavior patterns [Box and Jenkins, 1970].

6. Stationarity

When a time series appears to vary about some fixed level or mean, it is said to be stationary in the mean [Box and Jenkins, 1970]. Time series, as alluded to in the previous discussion on the decomposition method, may exhibit a long-term trend, be it positive or negative. In the case of the RTC Great Lakes graduation data, the observations fluctuate about the regression line in an upward-moving trend. This is called *non-stationary* behavior, and can be evidenced below:

RTC Great Lakes Graduates OCT94 - Jul98

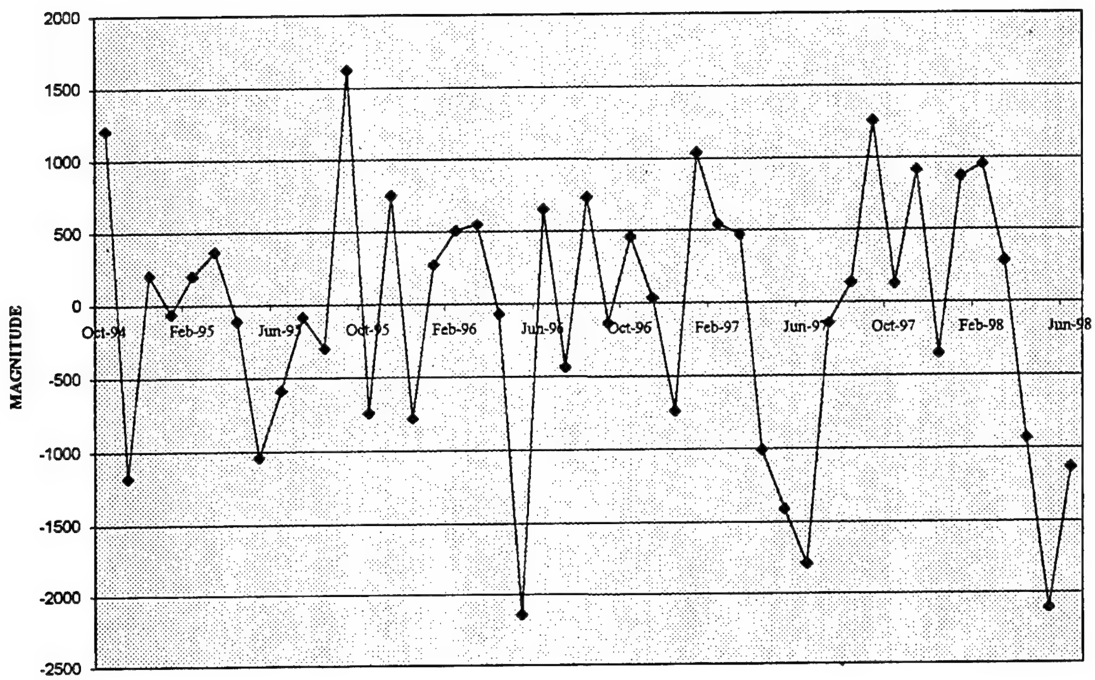


ARMA models apply to horizontal or stationary data distributions only [Box and Jenkins, 1970]. Fortunately, we can *difference*, or adjust, the original data to achieve stationarity [Bowerman and O'Connell, 1979]. In practice, the trend is removed by taking successive differences of the data to generate a new series. The following table indicates the result of taking one difference from the original data:

RTC GREAT LAKES DIFFERENCED DATA, OCT94 - JUL98					
Period	**Observation**	**Differenced Value**	**Period**	**Observation**	**Differenced Value**
	Z_t	$Z_t - Z_{t+1}$		Z_t	$Z_t - Z_{t+1}$
1	2951	1209	25	3588	456
2	1742	-1181	26	3132	31
3	2923	206	27	3101	-742
4	2717	-66	28	3843	1036
5	2783	198	29	2807	541
6	2585	365	30	2266	468
7	2220	-116	31	1798	-1008
8	2336	-1039	32	2806	-1413
9	3375	-587	33	4219	-1793
10	3962	-88	34	6012	-147
11	4050	-302	35	6159	132
12	4352	1625	36	6027	1250
13	2727	-740	37	4777	122
14	3467	749	38	4655	913
15	2718	-779	39	3742	-355
16	3497	269	40	4097	867
17	3228	504	41	3230	951
18	2724	546	42	2279	281
19	2178	-73	43	1998	-934
20	2251	-2140	44	2932	-2110
21	4391	653	45	5042	-1135
22	3738	-438	46	6177	6177
23	4176	731			
24	3445	-143			

A graph of the new series shows that the positive long-term trend has been removed from the data:

RTC Great Lakes Differenced Values, OCT94 - JUL98



If the *dth* difference of the original time series is stationary, a non-stationary data set may be represented by an ARMA model. This is referred to as an autoregressive-integrated-moving average model, ARIMA, of order (p,d,q) [Box and Jenkins, 1970]. In this case d=1, the first difference of the original data.

Mathematically, non-stationarity may be represented by a generalized autoregressive operator $\vartheta(B)$ [Box and Jenkins, 1970]:

$$\vartheta(B) = \Phi(B)(1-B)^d = \Theta(B)a_t \quad \text{,where} \tag{12}$$

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B_2^2 - \ldots - \Phi_p B^p$$
$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B_2^2 - \ldots - \Theta_q B^q$$

$\Phi(B)$ is the autoregressive operator of order p and is assumed to be stationary. $\Theta(B)$ is the moving average operator of order q. It is also convenient to consider an extension of the ARIMA model by adding a constant term Θ_0 [Box and Jenkins, 1970]:

$$\vartheta(B)Z_t = \Phi(B)(1-B)^d Z_t = \Theta_0 + \Theta(B)a_t \tag{13}$$

7. Model Selection

As mentioned earlier, p and q represent the order of the autoregressive and moving average processes, respectively. We can attempt to determine the order of these processes by means of a visual examination of the PACF and ACF graphs. In the PACF graph, the number of statistically significant partial autocorrelation coefficients is the same order as the AR(p) model, or p [Judge, et all, 1985]. In our case we see that there are at least two statistically significant coefficients, suggesting at least an AR(2) model. Similarly, the order of the MA(q) model, q, is determined from the ACF graph [Box and Jenkins, 1970]. Upon examination of the graph we find that the MA(3) model is a very likely candidate for consideration. With d=1, our best guess for the ARIMA model is the ARIMA(213) process. This can be easily verified with the MINITAB© software package:

ARIMA Model

Final Estimates of Parameters

Type		Estimate	St. Dev.	t-ratio
AR	1	1.1141	0.1323	8.42
AR	2	-0.8439	0.1446	-5.84
MA	1	1.3637	0.0234	58.27
MA	2	-1.0526	0.1613	-6.53
MA	3	0.7808	0.1514	5.16
Constant		21.6945	0.0133	1629.89

```
Differencing: 1 regular difference
No. of obs.:  Original series 46, after differencing 45
Residuals:    SS =  18396892  (backforecasts excluded)
              MS =    471715  DF = 39
```

The t-ratio is a measure of the standard error of each particular coefficient. It can be thought of as the number of standard errors from zero. For example, the AR(1) coefficient's t-ratio of 8.42 implies its significance is 8.42 standard errors from zero A high t-value indicates that the p and q coefficients play an increasingly important role in the model. Generally, t-ratios of over two are considered to be significant [Gujurati, 1995]. In our case, we see that the lowest t-ratio is of magnitude 5.16, and the highest of magnitude 1629.89, well over two and suggesting appropriate coefficients. Indeed, MINITAB© trials of other ARIMA processes such as ARIMA(212), ARIMA(211), and ARIMA(112), do not yield as good or consistent results as the ARIMA(213) model. It shall be our model henceforth.

We can derive an expression for the ARIMA(213) model. Using the autoregressive and moving average operators from Equation (13) and letting p=2,d=1,q=3, we have:

$$\left(1-\Phi_1 B-\Phi_2 B\right)(1-B)Z_t=\Theta_0+\left(1-\Theta_1 B-\Theta_2 B^2-\Theta_3 B^3\right)a_t \tag{14}$$

$$\left\{1-B-\Phi_1 B+\Phi_1 B^2-\Phi_2 B^2+\Phi_2 B^3\right\}Z_t=\Theta_0+\left(1-\Theta_1 B-\Theta_2 B^2-\Theta_3 B^3\right)a_t \tag{15}$$

$$\left\{1-(1+\Phi_1)B-(\Phi_2-\Phi_1)B^2-(-\Phi_2)B^3\right\}Z_t=\Theta_0+\left(1-\Theta_1 B-\Theta_2 B^2-\Theta_3 B^3\right)a_t \tag{16}$$

$$Z_t-(1+\Phi_1)Z_{t-1}-(\Phi_2-\Phi_1)Z_{t-2}-(-\Phi_2)Z_{t-3}=\Theta_0+a_t-\Theta_1 a_{t-1}-\Theta_2 a_{t-2}-\Theta_3 a_{t-3} \tag{17}$$

$$Z_t=(1+\Phi_1)Z_{t-1}-(\Phi_1-\Phi_2)Z_{t-2}-(\Phi_2)Z_{t-3}+\Theta_0+a_t-\Theta_1 a_{t-1}-\Theta_2 a_{t-2}-\Theta_3 a_{t-3} \tag{18}$$

Which is of the form

$$Z_t=\vartheta_1 Z_{t-1}-\vartheta_2 Z_{t-2}-\vartheta_3 Z_{t-3}+\Theta_0+a_t-\Theta_1 a_{t-1}-\Theta_2 a_{t-2}-\Theta_3 a_{t-3} \tag{19}$$

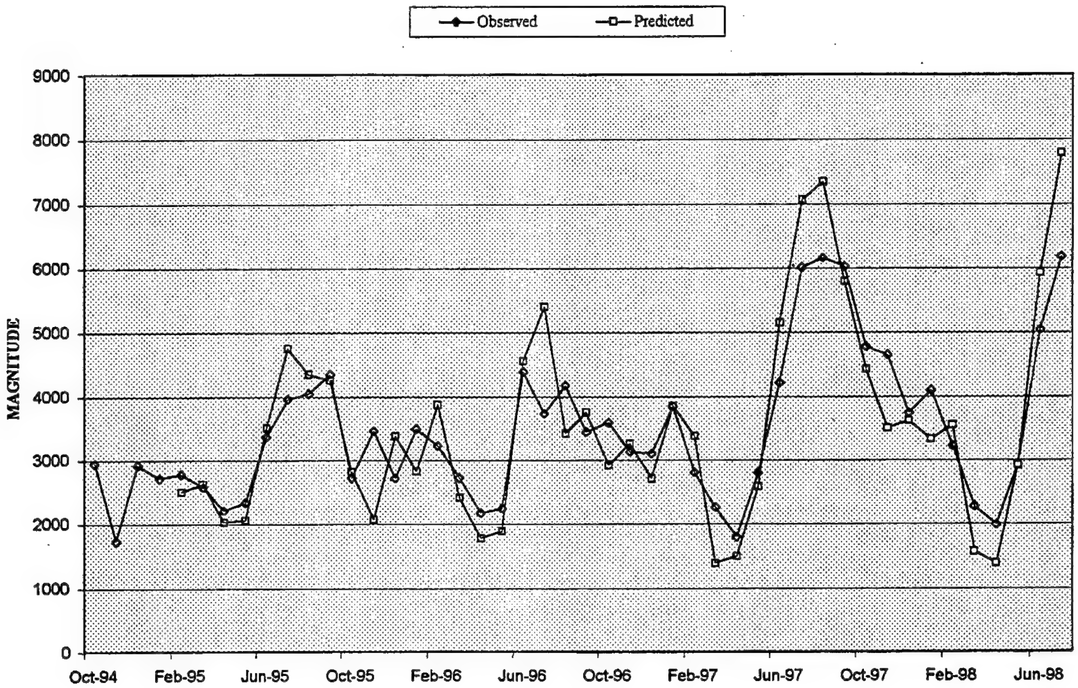
Substituting the ARIMA process coefficients found earlier, we obtain the mathematical expression for the RTC Great Lakes graduate data:

$$\begin{aligned} Z_t = {} & 2.1141\cdot Z_{t-1}-1.9580\cdot Z_{t-2}+0.8439\cdot Z_{t-3}+21.6945+a_t-1.3637\cdot a_{t-1} \\ & +1.0526\cdot a_{t-2}-0.7808\cdot a_{t-3} \end{aligned} \tag{20}$$

We can now utilize Equation (20) in spreadsheet form to obtain a tabular representation of the expression, as well as its graphical interpretation:

RTC GREAT LAKES GRADUATE DATA, ARIMA 213 PROCESS CALCULATION						
			(B)Z_t	Trend Line	Noise	Predicted
Period	Month	Observation	$Z_t - Z_{t-1}$	Estimation	$Z_{(t)}-Z^*$	Value
		Z_t		Z^*	a_t	Z_e
1	Oct-94	2951		2657	294	
2	Nov-94	1742	-1209	2693	-951	
3	Dec-94	2923	1181	2728	195	
4	Jan-95	2717	-206	2764	-47	
5	Feb-95	2783	66	2800	-17	2507
6	Mar-95	2585	-198	2836	-251	2623
7	Apr-95	2220	-365	2872	-652	2040
8	May-95	2336	116	2907	-571	2069
9	Jun-95	3375	1039	2943	432	3516
10	Jul-95	3962	587	2979	983	4758
11	Aug-95	4050	88	3015	1035	4356
12	Sep-95	4352	302	3051	1301	4262
13	Oct-95	2727	-1625	3086	-359	2824
14	Nov-95	3467	740	3122	345	2080
15	Dec-95	2718	-749	3158	-440	3380
16	Jan-96	3497	779	3194	303	2828
17	Feb-96	3228	-269	3230	-2	3871
18	Mar-96	2724	-504	3265	-541	2416
19	Apr-96	2178	-546	3301	-1123	1788
20	May-96	2251	73	3337	-1086	1894
21	Jun-96	4391	2140	3373	1018	4554
22	Jul-96	3738	-653	3409	329	5410
23	Aug-96	4176	438	3444	732	3428
24	Sep-96	3445	-731	3480	-35	3756
25	Oct-96	3588	143	3516	72	2916
26	Nov-96	3132	-456	3552	-420	3260
27	Dec-96	3101	-31	3588	-487	2714
28	Jan-97	3843	742	3623	220	3858
29	Feb-97	2807	-1036	3659	-852	3381
30	Mar-97	2266	-541	3695	-1429	1393
31	Apr-97	1798	-468	3731	-1933	1507
32	May-97	2806	1008	3767	-961	2591
33	Jun-97	4219	1413	3802	417	5154
34	Jul-97	6012	1793	3838	2174	7068
35	Aug-97	6159	147	3874	2285	7348
36	Sep-97	6027	-132	3910	2117	5795
37	Oct-97	4777	-1250	3946	831	4430
38	Nov-97	4655	-122	3981	674	3502
39	Dec-97	3742	-913	4017	-275	3624
40	Jan-98	4097	355	4053	44	3329
41	Feb-98	3230	-867	4089	-859	3550
42	Mar-98	2279	-951	4125	-1846	1573
43	Apr-98	1998	-281	4160	-2162	1389
44	May-98	2932	934	4196	-1264	2922
45	Jun-98	5042	2110	4232	810	5930
46	Jul-98	6177	1135	4268	1909	7789

Actual vs. ARIMA 213 Process with Constant



8. Forecasting

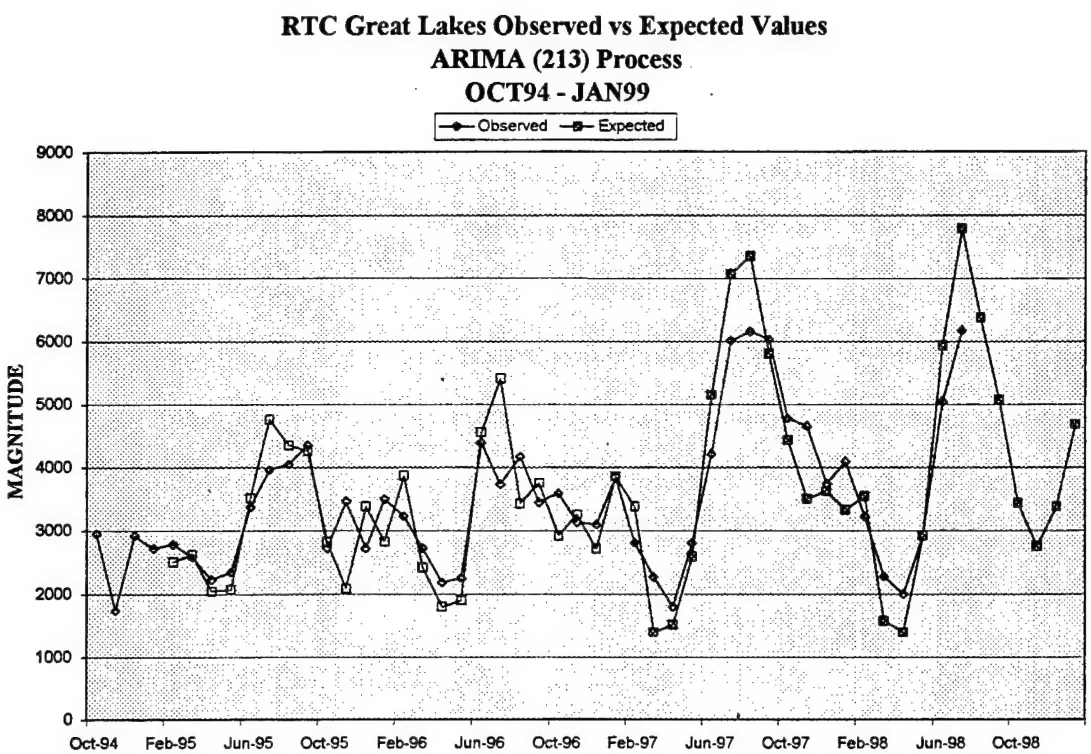
We perform a slight extension to Equation (20) in order to establish a forecast of future RTC Great Lakes graduation rates based on the ARIMA 213 process. To forecast a value for the variable Z, l values from origin t, we compute in spreadsheet form:

$$\tilde{Z}_{t+l} = 2.1141 \cdot Z_{t+l-1} - 1.9580 \cdot Z_{t+l-2} + 0.8439 \cdot Z_{t+l-3} + 21.6945 + a_{t+l} - 1.3637 \cdot a_{t+l-1} + 1.0526 \cdot a_{t+l-2} - 0.7808 \cdot a_{t+l-3} \quad (21)$$

Alternatively, we can allow the MINITAB© software package to generate the desired results:

RTC GREAT LAKES GRADUATE FORECAST				
Period	Month			Forecast Value
47	Aug-98			6376
48	Sep-98			5068
49	Oct-98			3439
50	Nov-98			2751
51	Dec-98			3380
52	Jan-99			4683

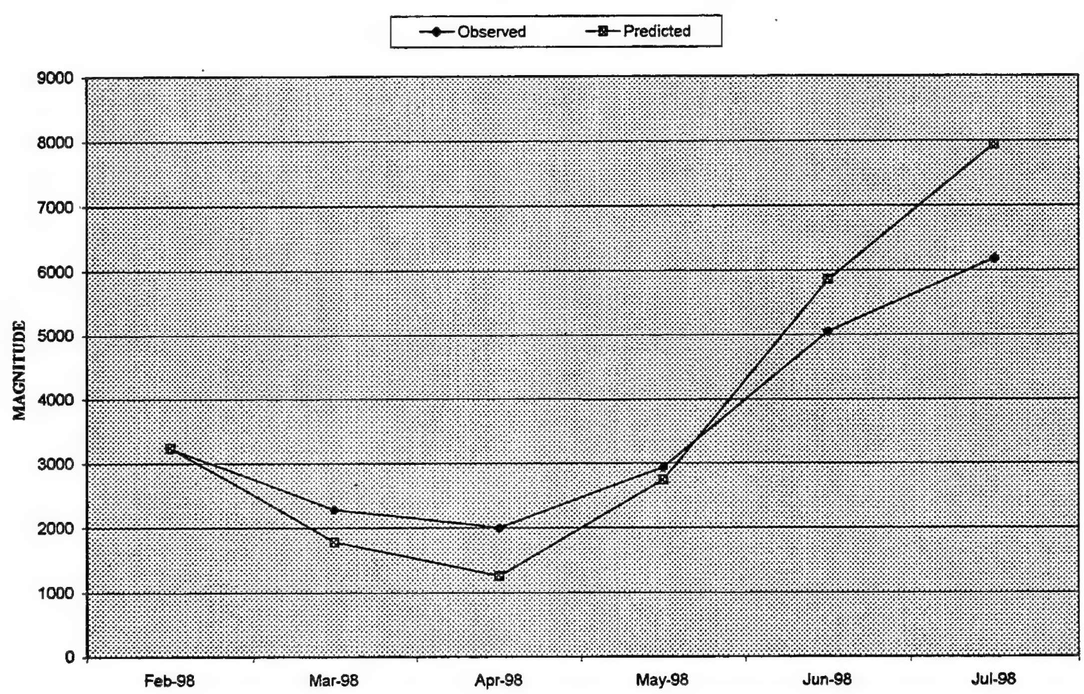
A graph of actual and expected values, to include the forecast figures, appears as follows:



9. Forecasting Error

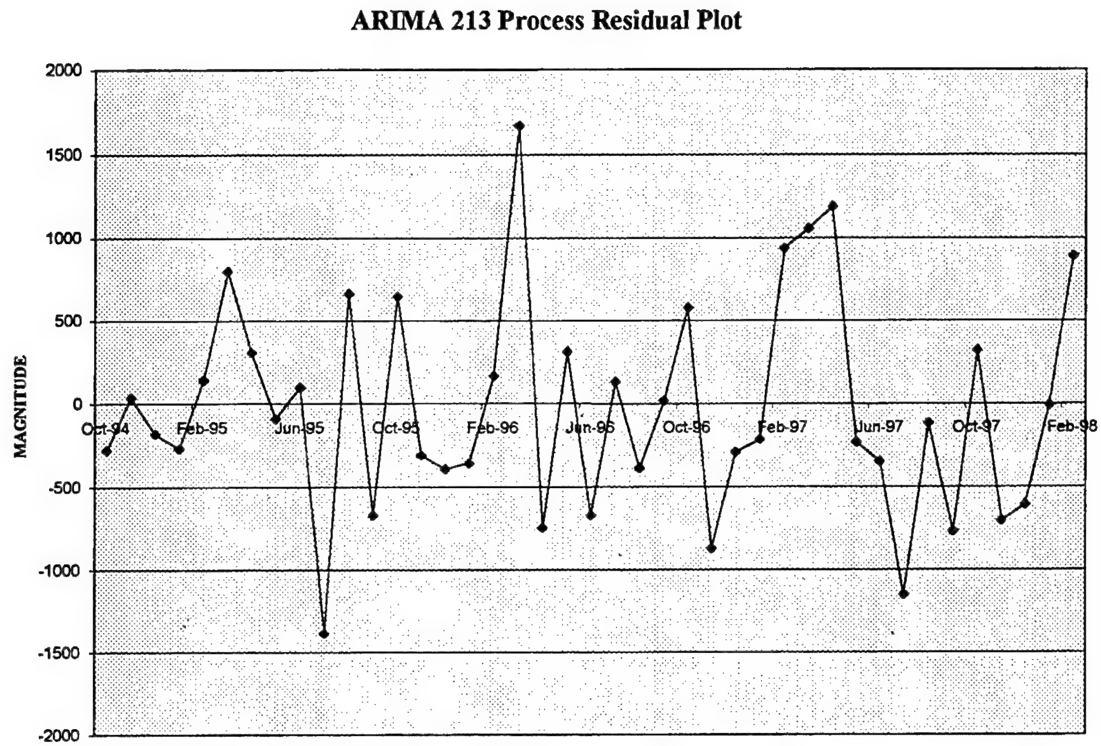
As with the multiplicative decomposition method, we back forecasted the results by performing a similar analysis with only the first forty observed values. As expected, different autoregressive and moving average coefficients were generated. A forecast was run with these particular values through period forty-six, and that forecast is contrasted to the actual observed values of periods forty to forty-six of the original observed values:

ARIMA 213 Process Back Forecasting Results



By visual inspection we observe that back forecasting provides an indication that the ARIMA(213) process is appropriate. The back forecasting results appear to model the actual observations. We soon consider the quantifiable error inherent in our model by way of the MSE calculation shortly.

The residual graph for the ARIMA 213 process which follows graphically indicates that all trend components of the data have been eliminated, leaving only a random component present [Bowerman and O'Connell, 1979]. No identifiable pattern could be found.



As before, we use the mean squared error of the forecasts in order to quantify the model's performance. Shown is the MSE for the forecasts of periods forty-one through forty-six, the back forecasted data:

ARIMA 213 PROCESS MSE CALCULATION						
		Observed	Forecast		Squared	
Period	Month	Value	Value	Error	Error	MSE
		Z_t	Z_e	$e_t = Z_t - Z_e$		
41	Feb-98	3230	3550	-320	102,400.00	726,484.17
42	Mar-98	2279	1573	706	498,436.00	
43	Apr-98	1998	1389	609	370,881.00	
44	May-98	2932	2922	10	100.00	
45	Jun-98	5042	5930	-888	788,544.00	
46	Jul-98	6177	7789	-1612	2,598,544.00	
				Sum	4,358,905.00	

We shall compare the ARIMA(213) process' MSE with the multiplicative decomposition method's MSE in the following section. For now we can assume by way of the back forecast and residual plot that the ARIMA(213) process is in itself a pertinent model which can be used in forecasting RTC Great Lakes' future graduation rates.

IV. DISCUSSION AND RECOMMENDATIONS

A. Discussion

We see from the back forecasting results that both the multiplicative decomposition method and ARIMA(213) process yield adequate results. We observe that the multiplicative decomposition method is the more conservative of the two. It underestimates the RTC Great Lakes recruit forecast, as opposed to the ARIMA(213) process' frequent overestimation. The multiplicative decomposition method's conservative numbers are also evidenced by its lower mean squared error figure. Lower MSE figures correspond to a better fit model [Bowerman and O'Connell, 1979]. For this reason one can conclude that the multiplicative decomposition method produces more accurate results. The multiplicative decomposition method's other allure is in its simplicity. Unlike the much more complicated ARIMA(213) process, the multiplicative decomposition method is built upon ratios and easily performed without special software or advanced mathematical knowledge.

On the other hand, one should not disregard the ARIMA(213) process' results altogether. For purposes of this study, the ARIMA(213) process has suffered from a low number of raw data observations. Ideally, the number of observed values should be at least approximately 50 [Box and Jenkins, 1970]. As the number of observations grow, the ARIMA(213) process should yield increasingly accurate results [Box and Jenkins, 1970].

B. Recommendations

It is recommended that the multiplicative decomposition and ARIMA analyses should be periodically performed and updated. As the number of observations under study increases, the parameters for both studies will surely change and ultimately lead to better models [Box and Jenkins, 1970]. In the short term, the multiplicative decomposition method should be employed. As more data becomes available, however, the ARIMA process should be reevaluated and considered.

C. Concluding Comments

The use of forecasting techniques can provide information to help alleviate many of the logistical problems at RTC Great Lakes and for the Navy. Knowledge of future months' recruit graduation rates can ease many of the effects of RTC's "summer surge." These unbalanced loads can be anticipated and prepared for not only by RTC, but also by follow-on schools, apprentice training, and manpower placement for the fleet.

The results seen here can be applied to a "feedback mechanism" which would be able to temper fluctuations and approximate the "level load" scenario, or constant output [Box and Jenkins, 1970]. This feedback mechanism is suggested for further study.

LIST OF REFFERENCES

Bowerman, B.L. and O'Connell, R., *Time Series and Forecasting; an Applied Approach*, Duxbury Press, 1979

Box, G.E.P. and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, Holden-Day, Inc., 1970

Brockwell, P.J. and Davis, R.A., *Time Series: Theory and Methods*, Springer-Verlag, Inc., 1991

Data Control Officer of the RTC, interview with the author, September 1998

Executive Officer of the RTC, interview with author, August 1998

Gujarati, D.N., *Basic Econometrics*, McGraw-Hill Inc., 1995

Harvey, A.C. *Time Series Models*, The MIT Press, 1993

Judge, G.G., Griffiths, W.E., Hill, R.C., Lukerpohl, H., and Lee, T.C. *The Theory and Practice of Econometrics*, John Wiley and Sons, 1985

Kennedy, P. A Guide to Econometrics, The MIT Press, 1979

http://www.ntcpao.com Naval Training Center Public Affairs Officer, RTC Website, July 1998

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Road, STE 0944
   Fort Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Road
   Monterey CA 93943-5101

3. Recruit Training Command, Great Lakes 2
   Commanding Officer
   Recruit Training Command
   3301 Indiana Street
   Great Lakes IL 60088-5300

4. Naval Center for Cost Analysis 1
   1111 Jefferson Davis Hwy
   Suite 400 West Tower
   Arlington, VA 22202-4306

5. Professor Kenneth J. Euske, Code SM/KU 1
   Naval Postgraduate School
   Monterey CA 93943

6. Professor Michael Cook, Code SM/MC 1
   Naval Postgraduate School
   Monterey CA 93943

7. LT Edward F. Bosque 1
   15524 Clark Ave
   Bellflower CA 90706